

A2  
cmcd.

technique for ranking documents within a database whose content has a large variation in quality and importance. Another aspect of the present invention is to provide a document ranking method that is scalable and can be applied to extremely large databases such as the world wide web. Additional aspects of the invention will become apparent in view of the following description and associated figures.--.

---

~~Please delete the paragraph at page 4, line 18.~~

---

A2  
cmcd.

Please ~~replace~~ the paragraph beginning at page 4, line 19, with the following rewritten paragraph:

---

--One aspect of the present invention is directed to taking advantage of the linked structure of a database to assign a rank to each document in the database, where the document rank is a measure of the importance of a document. Rather than determining relevance only from the intrinsic content of a document, or from the anchor text of backlinks to the document, a method consistent with the invention determines importance from the extrinsic relationships between documents. Intuitively, a document should be important (regardless of its content) if it is highly cited by other documents. Not all citations, however, are necessarily of equal significance. A citation from an important document is more important than a citation from a relatively unimportant document. Thus, the importance of a page, and hence the rank assigned to it, should depend not just on the number of citations it has, but on the importance of the citing documents as well. This implies a recursive definition of rank: the rank of a document is a function of the

A3  
encl.

~~replace the~~

--In one aspect of the invention, a computer implemented method is provided for scoring linked documents. The method includes identifying a plurality of documents, at least some of the documents being linked documents, at least some of the documents being linking documents, and at least some of the documents being both linked documents and linking documents, each of the linked documents being pointed to by a link in one or more of the linking documents; assigning a score to each of the linked documents based on scores associated with the one or more linking documents; and processing the linked documents according to the assigned scores.

In accordance with another implementation consistent with the present invention, a method for scoring linked documents includes receiving a search query from a user; identifying a plurality of documents responsive to the search query; locating incoming links to the identified documents from corresponding linking documents; assigning a score to each of the identified documents based on a number of the incoming links to the identified document and an importance of the incoming links; creating a ranked list based on the scores of the identified documents; and presenting to the user information about the identified documents in an order that is based on the ranked list.

In accordance with yet another implementation consistent with the present invention, a method for organizing linked nodes includes determining first link information for a linked node; determining second link information for linking nodes that

In accordance with a further implementation consistent with the present invention, a method of organizing linked documents includes: (a) identifying a first linked document; (b) identifying links between linking documents and the first linked document; (c) assigning a weight to each of the identified links; (d) determining a score for the first linked document based on (i) a number of the identified links between the linking documents and the first linked document, and (ii) the weights assigned to each of the identified links; (e) repeating steps (a) – (d) for a second linked document; and (f) organizing the first and second linked documents based on the determined scores.

Additional aspects, applications and advantages will become apparent in view of the following description and associated figures.--.

4

At page 6, line 17, please add the following paragraph:

A5

--Fig. 3 is a flowchart of one implementation of the invention.--.

Please replace the paragraph beginning at page 6, line 19, with the following rewritten paragraph:

--Although the following detailed description contains many specifics for the purposes of illustration, anyone of ordinary skill in the art will appreciate that many variations and alterations to the following details are within the scope of the invention. Accordingly, the following embodiments of the invention are set forth without any loss of generality to, and without imposing limitations upon, the claimed invention. For support in reducing the present invention to practice, the inventor acknowledges Sergey Brin, Scott Hassan, Rajeev Motwani, Alan Steremberg, and Terry Winograd.--.

At page 11, line 26, please add the following paragraph:

--Fig. 3 shows one embodiment of a computer implemented method for calculating an importance rank for N linked nodes of a linked database. At a step 101, an initial N-dimensional vector  $\mathbf{p}_0$  is selected. An approximation  $\mathbf{p}_n$  to a steady-state probability  $\mathbf{p}_\infty$  in accordance with the equation  $\mathbf{p}_n = \mathbf{A}^n \mathbf{p}_0$  is computed at a step 103. Matrix  $\mathbf{A}$  can be an NxN transition probability matrix having elements  $\mathbf{A}[i][j]$  representing a probability of moving from node i to node j. At a step 105, a rank  $r[k]$  for node k from a  $k^{\text{th}}$  component of  $\mathbf{p}_n$  is determined.--.

Please replace the paragraph beginning at page 11, line 27, with the following rewritten paragraph:

A 8  
--In one particular embodiment, a finite number of iterations are performed to approximate  $p_{\infty}$ . The initial distribution can be selected to be uniform or non-uniform. A uniform distribution would set each component of  $p_0$  equal to  $1/N$ . A non-uniform distribution, for example, can divide the initial probability among a few nodes which are known a priori to have relatively large importance. This non-uniform distribution decreases the number of iterations required to obtain a close approximation to  $p_{\infty}$  and also is one way to reduce the effect of artificially inflating relevance by adding unrelated terms.--.

Please replace the paragraph at page 12, line 6, with the following rewritten paragraph:

A 9  
cm T  
--In another particular embodiment, the transition matrix  $A$  is given by  
$$A = \alpha/N \mathbf{1} + (1-\alpha)B,$$
where  $\mathbf{1}$  is an  $N \times N$  matrix consisting of all 1s,  $\alpha$  is the probability that a surfer will jump randomly to any one of the  $N$  nodes, and  $B$  is a matrix whose elements  $B[i][j]$  are given by

$$B[i][j] = \begin{cases} 1/n_i & \text{if node } i \text{ points to node } j \\ 0 & \text{otherwise} \end{cases},$$

where  $n_i$  is the total number of forward links from node  $i$ . The  $(1-\alpha)$  factor acts as a damping factor that limits the extent to which a document's rank can be inherited by children documents. This models the fact that users typically jump to a different place in

~~the web after following a few links. The value of  $\alpha$  is typically around 15%. Including~~

~~this damping is important when many iterations are used to calculate the rank so that~~

~~there is no artificial concentration of rank importance within loops of the web.~~

~~Alternatively, one may set  $\alpha=0$  and only iterate a few times in the calculation.--~~

Please replace the paragraph beginning at page 12, line 27, with the following  
rewritten paragraph:

--Consistent with the present invention, there are several ways that this method can be adapted or altered for various purposes. As already mentioned above, rather than including the random linking probability  $\alpha$  equally among all nodes, it can be divided in various ways among all the sites by changing the  $\mathbf{I}$  matrix to another matrix. For example, it could be distributed so that a random jump takes the surfer to one of a few nodes that have a high importance, and will not take the surfer to any of the other nodes. This can be very effective in preventing deceptively tagged documents from receiving artificially inflated relevance. Alternatively, the random linking probability could be distributed so that random jumps do not happen from high importance nodes, and only happen from other nodes. This distribution would model a surfer who is more likely to make random jumps from unimportant sites and follow forward links from important sites. A modification to avoid drawing unwarranted attention to pages with artificially inflated relevance is to ignore local links between documents and only consider links between separate domains. Because the links from other sites to the document are not directly under the control of a typical web site designer, it is then difficult for the designer to artificially inflate the ranking. A simpler approach is to weight links from

A10  
Cmld

pages contained on the same web server less than links from other servers. Also, in addition to servers, internet domains and any general measure of the distance between links could be used to determine such a weighting.--.

---

Please ~~replace~~ the paragraph beginning at page 14, line 8, with the following rewritten paragraph:

---

A11  
Cmld

--Various implementations of the invention have the advantage that the convergence is very fast (a few hours using current processors) and it is much less expensive than building a full-text index. This speed allows the ranking to be customized or personalized for specific users. For example, a user's home page and/or bookmarks can be given a large initial importance, and/or a high probability of a random jump returning to it. This high rating essentially indicates to the system that the person's homepage and/or bookmarks does indeed contain subjects of importance that should be highly ranked. This procedure essentially trains the system to recognize pages related to the person's interests.--.

---

Please ~~replace~~ the paragraph beginning at page 15, line 10, with the following rewritten paragraph:

---

A12  
Cmld

--Another important application and embodiment of the present invention is directed to enhancing the quality of results from web search engines. In this application of the present invention, a ranking method according to the invention is integrated into a web search engine to produce results far superior to existing methods in quality and performance. A search engine employing a ranking method of the present invention

A12  
cmlld.

1. 1. The first part of the paper is devoted to a discussion of the  
 2. 2. importance of the problem of the  
 3. 3. determination of the  
 4. 4. of the  
 5. 5. of the  
 6. 6. of the  
 7. 7. of the  
 8. 8. of the  
 9. 9. of the  
 10. 10. of the  
 11. 11. of the  
 12. 12. of the  
 13. 13. of the  
 14. 14. of the  
 15. 15. of the  
 16. 16. of the  
 17. 17. of the  
 18. 18. of the  
 19. 19. of the  
 20. 20. of the  
 21. 21. of the  
 22. 22. of the  
 23. 23. of the  
 24. 24. of the  
 25. 25. of the  
 26. 26. of the  
 27. 27. of the  
 28. 28. of the  
 29. 29. of the  
 30. 30. of the  
 31. 31. of the  
 32. 32. of the  
 33. 33. of the  
 34. 34. of the  
 35. 35. of the  
 36. 36. of the  
 37. 37. of the  
 38. 38. of the  
 39. 39. of the  
 40. 40. of the  
 41. 41. of the  
 42. 42. of the  
 43. 43. of the  
 44. 44. of the  
 45. 45. of the  
 46. 46. of the  
 47. 47. of the  
 48. 48. of the  
 49. 49. of the  
 50. 50. of the  
 51. 51. of the  
 52. 52. of the  
 53. 53. of the  
 54. 54. of the  
 55. 55. of the  
 56. 56. of the  
 57. 57. of the  
 58. 58. of the  
 59. 59. of the  
 60. 60. of the  
 61. 61. of the  
 62. 62. of the  
 63. 63. of the  
 64. 64. of the  
 65. 65. of the  
 66. 66. of the  
 67. 67. of the  
 68. 68. of the  
 69. 69. of the  
 70. 70. of the  
 71. 71. of the  
 72. 72. of the  
 73. 73. of the  
 74. 74. of the  
 75. 75. of the  
 76. 76. of the  
 77. 77. of the  
 78. 78. of the  
 79. 79. of the  
 80. 80. of the  
 81. 81. of the  
 82. 82. of the  
 83. 83. of the  
 84. 84. of the  
 85. 85. of the  
 86. 86. of the  
 87. 87. of the  
 88. 88. of the  
 89. 89. of the  
 90. 90. of the  
 91. 91. of the  
 92. 92. of the  
 93. 93. of the  
 94. 94. of the  
 95. 95. of the  
 96. 96. of the  
 97. 97. of the  
 98. 98. of the  
 99. 99. of the  
 100. 100. of the

A13



Please replace the paragraph beginning at page 16, line 10, with the following

--Once a set of documents is identified that match the search terms, the list of

Please replace the paragraph beginning at page 16, line 18, with the following

--It will be clear to one skilled in the art that the above embodiments may be altered

IN THE CLAIMS:

Please cancel claims 1-8 without prejudice or disclaimer.

Please add new claims 9/36 as follows:

--9. A computer implemented method of scoring a plurality of linked

documents, comprising:

identifying a plurality of documents, at least some of the documents being linked